图书情报工作
第65卷 第12期 2021年6月

基于链路预测的协同药物组合推荐研究：面向疾病并发症诊疗*

■ 雷鸣[1] 夏梦鸽[1] 汪雪锋[1] 刘佳[2]
[1] 北京理工大学管理与经济学院 北京 100081
[2] 中国传媒大学媒体融合与传播国家重点实验室 北京 100024

摘 要：[目的/意义]药物组合相比单一药物在临床治疗中存在多种优势，但药物数量的快速增长为药物组合筛选和推荐带来挑战，因此设计有效的预测方法为药物研发人员推荐更易产生协同作用的药物组合从而提高筛选效率具有重要意义。[方法/过程]面向疾病并发症诊疗需求，基于链路预测构建协同药物组合推荐模型，首先利用SAO语义挖掘识别医学文献中的并发症信息，在此基础上利用医学数据库构建“疾病－药物－靶点”异质网络，并引入链路预测方法对网络进行药物作用机制的相似性评估，预测哪些药物组合更可能产生协同作用，进而依据预测结果针对某个疾病或某对并发症进行药物组合推荐。[结果/结论]肠道疾病数据实证分析结果表明协同药物组合预测模型具有有效性和实用性。

关键词：SAO语义挖掘 链路预测 社会网络分析 药物组合推荐 疾病并发症
分类号：G250
DOI：10.13266/j.issn.0252－3116.2021.12.012

1 引言

老药新用是目前重要的制药策略之一，科技部生物中心强调新冠肺炎药物的研发要遵循该策略，因其可以提高研发效率，同时可以保证安全性，而利用筛选试验研制药物组合是其中一种主要方向。协同药物组合指的是两种以上的药物合用时，如果它们的作用方向一致，能够实现彼此增强的效果，即组合总效应超过各药单用时效应的总和[1]。协同药物组合既可以是治疗一种疾病的药物组合，如治疗艾滋病的抗病毒组合，也可以是治疗多种疾病的药物组合，如氨氯地平阿伐他汀片，该药物的两种成分苯磺酸氨氯地平和阿伐他汀丐，前者用于治疗高血压，后者用于治疗高胆固醇。协同药物组合是在中西医结合背景下提出的，既能体现中药的整体协调性，又能体现西药作用机制靶点的明确性[2]。在临床治疗中，协同药物组合相比于单一药物具有提高疗效、降低毒副作用和降低抗药性等优势[3]。协同药物组合来源于对大量随机药物组合进行的筛选实验，但随着单药数量的快速增加，随机组合数量暴增，筛选效率降低，因此有必要开发有效的协同药物组合推荐方法为筛选试验缩小范围从而提高药物组合的研发效率。

目前，协同药物组合推荐主要基于三类方法：高通量筛选法、基于计算的预测模型和基于网络的预测模型。其中高通量筛选协同药物组合具有过高的实验要求，另外两种方法则随着数学、计算机科学和生物信息学的发展不断趋于融合，结合计算与网络生物学的预测方法是近年来的研究热点。例如，J. Zou等提出了一个基于网络的机器学习模型，通过整合药物－药物组合作用、药物－靶标相互作用和靶标－蛋白相互作用，重新构建药物组合网络，利用网络拓扑性质和功能性质构建特征向量，采用支持向量机工具预测协同药物组合[4]。S. Alaimo等通过综合表型、结构和基因组相似性，在药物相互作用网络的基础上，使用机器学习方法进行药物相互作用预测[5]。K. Pang等开发了一种算法，通过从Drugbank数据库构建药物－靶点相互作

* 本文系国家自然科学基金面上项目“生物医学领域潜在颠覆性技术识别方法研究”（项目编号：72074020）研究成果之一。
作者简介：雷鸣（ORCID：0000－0003－1746－4090），博士研究生；夏梦鸽（ORCID：0000－0003－0612－5659），硕士研究生；汪雪锋（ORCID：0000－0002－4857－6944），教授，博士，通讯作者，E-mail：wxf5122@bit.edu.cn；刘佳（ORCID：0000－0003－2627－933X），副教授，在站博士后。
收稿日期：2020－07－31 修回日期：2021－02－11 本文起止页码：122－129 本文责任编辑：易飞

雷鸣，夏梦鸽，汪雪锋，等.基于链路预测的协同药物组合推荐研究：面向疾病并发症诊疗[J].图书情报工作，2021，65(12)：122－129.

用网络来识别协同药物组合，利用药物综合相似性信息来预测药物相互作用[6]。这些模型以已有研究发现为理论依据，即药物的作用机理越相似，其协同作用概率越大。但这些研究缺乏对疾病并发症状的考虑，并发症即为共发生率高的疾病，治疗并发症是研发协同药物组合的主要目的之一，因此本文引入文本挖掘和社会网络方法对此进行改进。

链路预测是指利用已知网络信息，预测网络中目前尚未产生连边的两个节点间未来产生连接的可能性[7]。该方法假设两个节点越相似，它们之间发生关系的可能性就越高，就越容易产生连接。链路预测广泛应用于蛋白质互作网络、代谢网络、疾病关联网络中，以识别相关度高的实体。因其广泛适用性，链路预测已在计算机领域、社交网络和生物分子网络中积累了广泛的研究基础和成果，特别是在生物分子网络中，蛋白质－蛋白质相互作用网络、代谢网络和食物网络等生物网络的应用最为广泛。这是因为在实验室或领域中发现联系或相互作用的代价是昂贵的，而高精度的预测可以降低实验成本，加快揭示真相的速度[8]。以蛋白质交互网络为例，由于蛋白质间大部分的相互作用依然未知，在已知网络拓扑结构的基础上提前设计精确的链路预测方法，有利于提高实验的成功率并节省实验消耗成本[9]。例如，贾成龙提出基于网络模型的长非编码 RNA(lnc RNA)－疾病关联预测方法，整合 lnc RNA－疾病关联数据、lnc RNA 相似性网络和疾病相似性网络从而得到 lnc RNA－疾病异质网络，然后基于链路预测算法，对 lnc RNA 和疾病关联关系进行预测[10]。因此，本文基于链路预测构建协同药物组合推荐模型。首先利用 SAO(Subject-Action-Object)语义挖掘识别出医学文献中的并发疾病信息对，并利用数据库中的药物信息在并发疾病网基础上构建“疾病－药物－靶点”异质网络，最后利用链路预测方法评估不同药物作用机制的相似性，预测哪些药物组合更可能产生协同作用。为验证模型的有效性，本文以肠道疾病为例，对协同药物组合进行了预测及推荐，研究表明该推荐模型可提供筛选实验的优先级，进而提高药物组合的研发效率。

2 方法框架

本文通过提取医学文献中的 SAO 语义结构构建并发疾病网络，基于药物数据库根据这些疾病名称构建“疾病－药物－靶点”异质网络，并利用链路预测算法评估不同药物间作用机理的相似度，预测治疗某对并发症或单一症状的潜在协同药物组合，为药物研发人员推荐药物组合筛选实验的优先级。本文构建协同药物组合推荐模型的具体流程如图 1 所示：

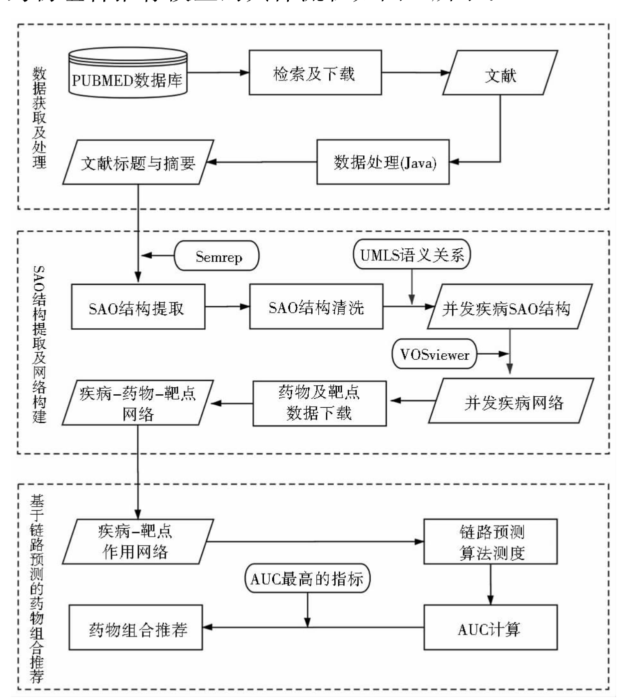


图 1　协同药物组合推荐模型构建流程

2.1 数据获取与预处理

医学文献中包含大量疾病间关联信息，且文献相比电子病历具有更高的专业性，因此本文基于 PubMed 医学文献数据库进行并发症挖掘和协同药物组合推荐。由于论文核心内容及研究成果一般出现在标题或摘要中，因此本文首先利用 Java 代码进行数据抽取，仅保留文献标识号、标题和摘要等信息，作为后续并发症识别的数据基础。

2.2 SAO 结构提取及并发网络构建

2.2.1 SAO 结构提取

SAO 是一种从文本语料中抽取出的主谓宾三元组结构[11]。其中，S 和 O 代表系统组件，A 是用来表达 Subject 和 Object 间关系的动词[12]。SAO 语义挖掘常被应用于科技情报信息提取[13]和专利相似性分析[14]等。由于该方法能保证两个关键词在同一句话中，故相比普通的关键词共现对关联关系的挖掘具有更高的准确性，对医学文本信息提取更具价值。而 Semrep 软件拥有 UMLS 的超级叙词表、语义网络和专家词典的强大功能，因此提取出的 SAO 结构拥有医学专业性[15]，故本文借助 Semrep 工具提取 SAO 结构，其表达示例如下：

23974065. ab. 1 | relation | C0014335 | Enteritis | dsyn |

dsyn|||COEXISTS_WITH|C1290894|Chronic disease of immune system|dsyn|dsyn||

其中,"Enteritis"和"Chronic disease of immune system"是从摘要第一句话中提取出的医学实体,"COEXISTS_WITH"表示这两个词之间的语义关系为"共存","dsyn"表示该实体为"疾病或症状"。由于本文旨在识别并发症,因此仅保留实体语义类型均为"dsyn"的 SAO 结构。

2.2.2 并发症信息识别

考虑到 Semrep 软件中界定的语义关系中"COMPLICATES""ASSOCIATED_WITH""CAUSES""AFFECTS""PREDISPOSES""MANIFESTATION_OF""PRECEDES""COEXISTS_WITH"均可代表疾病间具有并发关系[16],因此本文仅筛选包含这些语义关系的 SAO 结构。利用筛选出的具有并发关系的疾病对构建并发疾病网络,连边权重为不同疾病的共现频次,并利用 VOSviewer 软件形成可视化网络图谱,根据该网络则可以探索某疾病的并发症状况,具体构建流程如图 2 所示:

$S\text{-}O_1$	疾病1	疾病2
$S\text{-}O_2$	疾病1	疾病3
$S\text{-}O_3$	疾病3	疾病4
$S\text{-}O_4$	疾病2	疾病5
$S\text{-}O_5$	疾病5	疾病6
$S\text{-}O_6$	疾病6	疾病7
$S\text{-}O_7$	疾病6	疾病5

→

	疾病1	疾病2	疾病3	疾病4	疾病5	疾病6	疾病7
疾病1	0	1	1	0	0	0	0
疾病2	1	0	0	0	1	0	0
疾病3	1	0	0	1	0	0	0
疾病4	0	0	1	0	0	0	0
疾病5	0	1	0	0	0	2	0
疾病6	0	0	0	0	2	0	1
疾病7	0	0	0	0	0	1	1

→

图 2　并发疾病网络构建流程

2.3 疾病－药物－靶点异质网络构建

在并发疾病网络基础上,为进行协同药物组合推荐,需要明确每个疾病对应的药物。CTD 医学数据库(Comparative Toxicogenomics Database)中的疾病名称大多来源于 PubMed 文献数据库,因此选其作为药物查询来源可有效避免 CTD 医学数据库与 PubMed 文献疾病名称不匹配的问题。将"疾病－药物"对应关系转化为共现矩阵,利用 VOSviewer 进行可视化,构建如图 3 所示的"疾病－药物"网络:

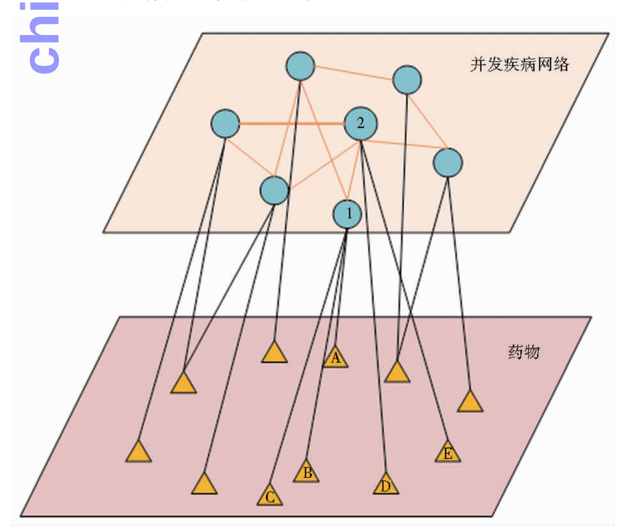


图 3　疾病－药物网络示意

以往的药物组合研究通常以治疗某个单独疾病为目的,如疾病 1。为预测协同药物组合,则需对所有能够治疗疾病 1 的药物进行随机组合并一一筛选,如图 3 中的三种药物 A、B、C。考虑到三者或三者以上药物配伍使用的机理较为复杂,目前的药物组合预测研究大多以两种药物间的组合为主,如图 3 所示例的疾病 1 则有可能形成三种随机药物组合:A＋B、A＋C、B＋C,药物组合预测研究则主要围绕判定这三个组合中哪些更可能作为协同药物组合展开[17]。

由于研究发现药物协同作用概率与作用机理相似度成正比[18],因此可通过评估药物作用机制相似度来预测协同药物组合。有学者研究指出,药物组合作用机理本质就是药物靶标在生物系统中的协同调控[19－20]。靶点的相关性在很大程度上决定药物作用机理的相似性,即评估两个药物能否形成协同组合,首先需要判断两者作用靶点是否相似或相关。J. Zou 等通过研究协同药物组合及其网络拓扑关系发现协同药物组合靶点往往存在于邻域,瞄准不相关的靶点往往不具有协同作用[4]。B. Rita 等研究也发现拮抗作用几乎均发生在靶点不同的药物之间,协同作用更多发生在相同靶点的药物中[21]。因此本文以靶点作为药物作用机理的特征信息,即用药物靶点相关性表征作用机制相似性,进一步判断哪些组合更可能是协同药物组合。利用 Drugbank 数据库查找药物的靶点,在"疾病－药物"网络基础上构建"疾病－药物－靶点"异质网络,如图 4 所示:

雷鸣，夏梦鸽，汪雪锋，等.基于链路预测的协同药物组合推荐研究：面向疾病并发症诊疗[J].图书情报工作，2021，65(12)：122－129.

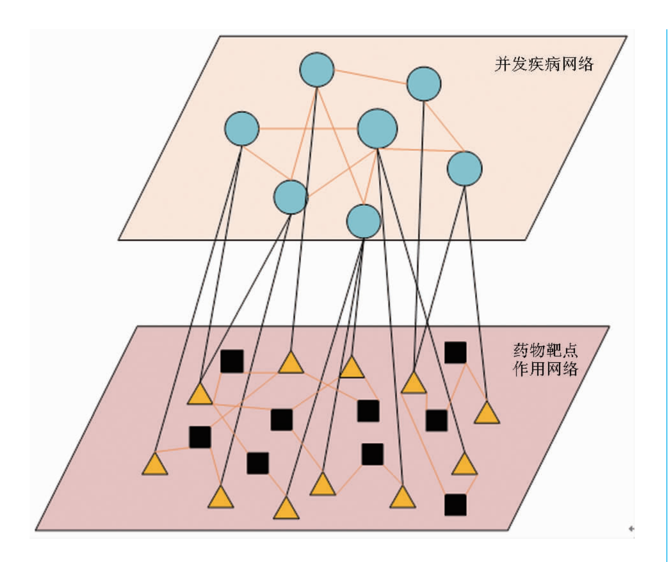


图 4　疾病－药物－靶点网络示意

2.4　协同药物组合推荐

2.4.1　链路预测指标选取

在疾病－药物－靶点异质网络中的药物靶点作用网络基础上，利用链路预测评估药物相似性。基于网络结构信息定义的相似性称为结构相似性，本文选取10个具有代表性的基于结构相似性的链路预测指标分别测度网络，表1列出了这些指标及计算公式。

应用上述指标对网络进行测度，可以计算出两两节点间的相似度，每对节点被赋予一个分数值。不同指标计算结果不同，为此还需评估指标的测度准确性。目前对网络结构特征和算法性能之间关系尚未开展深入研究，所以链路预测均通过准确性验证来选择对特定网络预测性能好的指标。

表 1　链路预测指标及计算公式

指标	计算公式	主要变量含义
CN(共同邻居数)	$CN_{xy}=\|\varphi(x)\cap\varphi(y)\|$	$\varphi(x)$表示节点 x 的邻居集合
Salton(余弦相似度)	$S_{xy}=\frac{\|\varphi(x)\cap\varphi(y)\|}{\sqrt{k_x*k_y}}$	$\varphi(y)$表示节点 y 的邻居集合 k_x 和 k_y 分别表示节点 x、y 的度
Jaccard	$J_{xy}=\frac{\|\varphi(x)\cap\varphi(y)\|}{\|\varphi(x)\cup\varphi(y)\|}$	
AA(*Adamic-Adar* 指标)	$AA_{xy}=\sum_{z\in\varphi(x)\cap\varphi(y)}*1/log\ k(z)$	
PA(优先链接指标)	$PA_{xy}=k_x*k_y$	
LP(局部路径指标)	$LP_{xy}=(A^2)_{XY}+\alpha(A^3)_{xy}$	A 表示网络的邻接矩阵，$(A^3)_{xy}$表示节点 x 和节点 y 之间长度为3的路径数目，α 是可调参数
Katz(全局路径指标)	$Katz_{xy}=\sum\sum_{l=1}^{\infty}\beta^l\cdot\|pathspaths_{xy}^{(l)}\|$ $=\beta A+\beta^2A^2+\beta^3A^3+\cdots$	β 是可调参数 $paths_{xy}^{(l)}$为连接节点的路径中长度为 L 的路径数
ATC(平均通勤时间)	$n(x,y)=M(l_{xx}^{+}+l_{yy}^{+}-2l_{xy}^{+})$	l_{xy}^{+}为矩阵 l^{+}中相应位置的元素
LRW(局部随机游走)	$LRW_{xy}=q_x*\pi_{xy}(t)+q_y*\pi_{yx}(t)$	P 为网络的马尔科夫概率转移矩阵，设各节点的初始资源分布为 q_x，即点 x 的度占网络中全部连边的比例
SRW(叠加的局部随机游走)	$SRW_{xy}=\sum_{l=1}^{t}LRW_{xy}(l)$	

AUC是常用的准确性评估指标[22]，具体评估过程如下：将网络现有边随机分为训练集和测试集，数据集大小比例为9∶1，将节点间尚未连接的边称为不存在的边，如果一个指标的计算结果中能尽量使测试集边的分数大于不存在的边，那么其预测效果越好。按照这种思想，每次随机地从测试集中选一条边，再随机选择一条不存在的边，比较两条边的相似度分值。若测试集的边的分数值大于不存在的边的分数，记为1分；如果两者分数相等，记为0.5分，否则记为0分。这样独立比较n次，设有 n_1 次测试集中的边的分值大于不存在的边的分值，有 n_2 次两分数相等，则AUC定义为：

$$\text{AUC}=\frac{n_1+0.5n_2}{n} \qquad (1)$$

将AUC值最大指标的计算结果作为药物间相似度值。由此可见，链路预测方法是从众多指标中筛选最适合特定网络的指标进行相似度计算。以往药物相似性计算均利用余弦相似度进行药物的靶点重合度测度，忽略了药物存在不相同却相关靶点的情况。由表1可知链路预测指标本身包含了余弦相似度指标，如果在预测结果中余弦相似度指标的AUC准确值最高，则对该网络使用链路预测与余弦相似度实际为同一结果；如果余弦相似度指标的AUC准确值不是最高，则证明链路预测方法相对于余弦相似度方法更适用于该网络。

2.4.2　协同药物组合推荐

基于链路预测结果的协同药物组合推荐示意如图5所示，其中左图即为基于药物靶点作用网络的链路

预测结果。在计算结果中删除相似值为零的节点连边，剩余即为链路预测新产生的连边，如图 5 中 A 和 E、B 和 C、B 和 D 间的连边。将链路预测结果添加到“疾病－药物”网络中，如图 5 的右侧部分，在此基础上，针对每一种疾病或者每一对并发症，都可以给出推荐的药物组合实验筛选顺序。

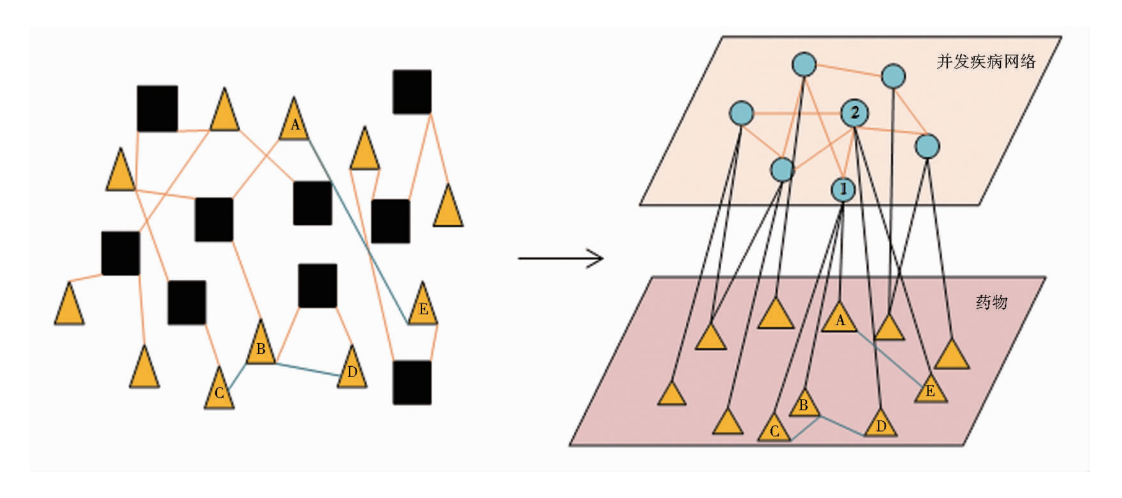


图 5　基于链路预测的协同药物组合推荐示意

由于药物组合是在中西医结合背景下提出的，中药本身即为治疗多种并发症状的协同药物组合，因此可根据中药的配伍过程进行协同药物组合推荐。如疾病 1 和疾病 2 是一对并发症，则治疗这对并发症的药物随机组合有 6 个：A＋D、A＋E、B＋D、B＋E、C＋D、C＋E。若链路预测结果显示在药物 A 与 E、B 与 D 之间有连边，则其为协同药物组合的概率就更大，可以在协同药物组合实验筛选中优先验证这两个组合。在治疗单个疾病时，如前文所述治疗疾病 1 的三种药物，随机组合中的 B 和 C 有连边，则可优先验证这个组合。

3　实证研究

考虑到肠道疾病具有较强复杂性，对并发症治疗有着重要需求，且治疗肠道疾病的药物种类繁多，协同药物组合筛选存在困难，因此本文以肠道疾病为例验证推荐模型有效性。在 PubMed 数据库以(“intestinal diseases”[MeSH Terms] OR (“intestinal”[All Fields] AND “diseases”[All Fields]) OR “intestinal diseases”[All Fields]) AND ((“1900/01/01”[PDAT]:“2019/07/29”[PDAT]) AND “humans”[MeSH Terms] AND English[lang])为检索式，检出并下载肠道疾病医学文献 422 621 篇(检索时间为 2020 年 1 月 20 日)，并通过 Java 代码抽取标题、摘要、文献标识号等信息。

利用 Semrep 提取出 SAO 结构 2 336 540 条，其中 S 和 O 的语义类型均为“dsyn”(表示该实体为“疾病或症状”)的 SAO 结构记录共 14 426 条。为挖掘并发症，从中筛选出表征并发疾病关系的有效记录为 1 221 条，将其转化为共现矩阵并输入到 VOSviewer 软件中，形成如图 6 所示的并发疾病网络图：

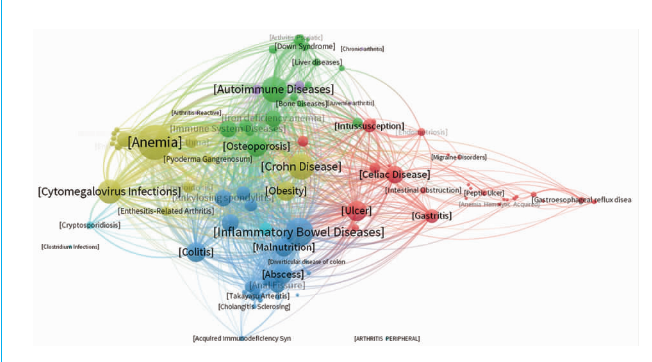


图 6　肠道疾病的并发疾病网络

针对其中共现频次高的并发疾病对进行分析，发现其发病机制间的确存在关联，如隐孢子虫病和艾滋病的传播途径有着相似之处[23]；乳糜泻与炎症性肠病都属于自身免疫性疾病，且都和肠道菌群失调有关[24]，拥有共同的易感基因[25]；克罗恩病会导致维生素 K 水平降低，而骨质与钙的结合需要维生素 K 的作用，因此克罗恩病患者会有很大概率罹患骨质疏松。这证明基于 SAO 语义挖掘的并发症识别具有效性。

在并发疾病网络基础上，利用 CTD 医学数据库根据并发症名称查找药物，构建肠部疾病的“疾病－药物”网络，然后利用 Drugbank 数据库查找这些药物的靶点，构建“疾病－药物－靶点”异质网络，具体如图 7、图 8 所示，为有效区分节点类别，在图中以中括号标识疾病节点。

基于其中的药物靶点作用网络进行药物组合预

雷鸣，夏梦鸽，汪雪锋，等.基于链路预测的协同药物组合推荐研究：面向疾病并发症诊疗[J].图书情报工作，2021，65(12)：122－129.

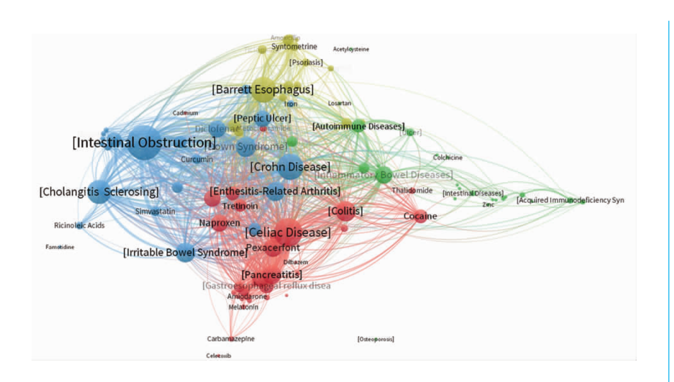


图7　肠道疾病－药物网络

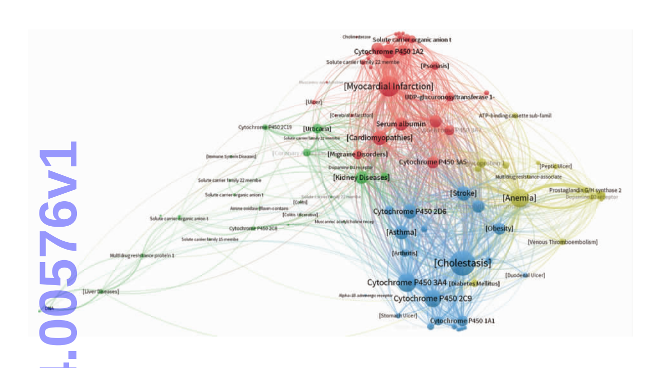


图8　肠道疾病－药物－靶点异质网络(局部)

测，通过链路预测算法测度药物相似度。其中链路预测指标计算节点相似性以及AUC指标计算均通过Python语言实现，算法输入为网络所有节点编号及网络现有连边，算法输出为基于各种指标计算得到的节点间相似度矩阵以及每一个指标的准确度AUC值(见表2)。从表2可以看出，Katz指标预测效果最好。由表1可知Katz指标考虑了网络中的全部路径，包括药物间三阶以上共同邻居；而使用余弦相似度方法则相当于链路预测里的Salton指标，其准确度低于0.5，这说明对于药物靶点作用网络应用链路预测方法比余弦相似度效果更好。

表2　链路预测指标AUC值

链路预测指标	AUC值	链路预测指标	AUC值
Katz	0.956 846	ATC	0.898 134
LP	0.923 104	Salton	0.483 125
PA	0.919 334	CN	0.483 122
SRW	0.918 104	Jaccard	0.483 000
LRW	0.917 472	AA	0.482 945

如果药物间相似度为0，则基本可以认为其作用机理不同，协同作用概率极低，则删除这一对组合，由此根据链路预测结果得到协同药物组合，接下来可针对具体疾病进行协同药物组合推荐。例如，有18种治疗子宫内膜异位的药物，则药物的随机组合有153种，其中在预测结果中有6对药物产生连边，即相似度不为0。这6对药物组合分别为：Rofecoxib与Tibolone、Dinoprost与Sulindac、Danazol与Progesterone、Mifepristone与Tibolone、Elagolix与Leuprolide、Leuprolide与Sulindac，可将这些作为协同药物组合推荐给药物研发人员进行优先验证，以期有效降低研究工作量。为验证模型预测结果的准确性，本文通过查找药物组合的相关资料进行分析。例如，查找治疗子宫内膜异位症的相关研究发现，Drugbank数据库包含的药物相互作用记录显示，药物对“Rofecoxib”(罗非昔布)与“Tibolone”(替勃龙)的确具有协同作用，“Rofecoxib”的使用会增加“Tibolone”药效活性，而这个药物组合位列预测结果第一位。另外，“Lynestrenol与Simvastatin”“Rofecoxib与Simvastatin”“Zafirlukast与Naproxen”“Celecoxib与Progesterone”“Ibuprofen与Elagolix”“Celecoxib与Naproxen”“Danazol与Mifepristone”“Ibuprofen与Simvastatin”等药物对则存在明显的拮抗效果，但这些药物对均不包含在药物组合预测结果中。

在面向并发症治疗时，同样可结合预测结果进行协同药物组合推荐。例如，由图6可知“Cholangitis Sclerosing”(硬化性胆管炎)和“Pancreatitis”(胰腺炎)是并发程度较高的疾病对，对研发协同药物组合同时治疗两种症状的需求较高。而治疗硬化性胆管炎和胰腺炎的药物随机组合有32种，其中有8个组合在链路预测结果中产生连边，可将其推荐给药物研发人员进行优先验证。同样，我们通过查找相关资料发现，美国在2018年上市了一种组合性质的药物“Festal Plus”，它可以同时减缓硬化性胆管炎和胰腺炎两种疾病，该药物的主要成分为Ursodeoxycholic acid(熊去氧胆酸)和Pancrelipase(胰脂酶)，前者主要用于治疗胆管炎，后者可用于治疗因囊性纤维化，慢性胰腺炎引起的胰腺功能不全。这证明同时使用Ursodeoxycholic acid和Pancrelipase可以产生协同作用，且已被研发成为组合药物，而熊去氧胆酸与胰脂酶包含在链路预测结果中，这说明本文构建的协同药物组合推荐模型可以给予具有协同效应的药物对更高的优先级，即协同药物组合推荐模型具有有效性。

4　结语

科技部生物中心在介绍新冠肺炎药物疫苗研发新情况时强调，科研攻关组在药物研发过程中要遵循老

药新用的研发策略。协同药物组合符合老药新用的制药策略,且于临床治疗有着多种优势,尤其针对复杂疾病的治疗,更有降低并发症患者用药复杂性的优势,设计有效预测方法提高药物组合筛选的效率尤为重要。本文借助网络生物学思想,结合数学模型和计算机技术,利用 SAO 语义挖掘技术和社会网络分析方法从并发症诊疗视角出发,构建了以靶点表征药物作用机理的协同药物组合预测框架,并引入链路预测方法对药物相似度计算做出改进,有效弥补了以往研究中忽略药物具有不相同但相关靶点的情况,实证研究发现对于药物靶点作用网络,链路预测算法相比余弦相似度具有更高的准确性。同时,通过查找相关医学资料,验证了部分预测结果,说明本文模型具有有效性和实用性,有助于为推荐给药物研发人员进行协同药物组合优先验证提供保障,从而有效降低研究者工作量。

本文研究不足之处在于:受学科专业限制,尚无法通过医学实验全面验证模型所有预测结果的准确性,但已通过现有资料验证了部分预测结果的准确性。同时,本文仅使用靶点信息度量药物相似性,药物作用机制信息相对单一,未来可尝试利用多维信息(比如药物化学式、副作用、组成成分等)测算药物相似性以及基于医学文本全文提取 SAO 结构以进一步提升预测效果。

参考文献:

[1] SUN X, VILAR S, TATONETTI N P. High-throughput methods for combinatorial drug discovery[J]. Science translational medicine, 2013, 5(205):205rv1.

[2] 陈畅, 马兴, 满国峰, 等. 组合药物——现代药物研发的新趋势[J]. 中国实验方剂学杂志, 2012, 18(14):310-312.

[3] SUCHER N J. Searching for synergy in silico, in vitro and in vivo[J]. Synergy, 2014, 1(1):30-43.

[4] ZOU J, JI P, ZHAO Y L, et al. Neighbor communities in drug combination networks characterize synergistic effect[J]. Molecular biosystems, 2012, 8(12):3185-3196.

[5] ALAIMO S, BONNICI V, CANCEMI D, et al. DT-Web: a Web-based application for drug-target interaction and drug combination prediction through domain-tuned network-based inference[J]. BMC system biology, 2015,9(S3): S4-S4.

[6] PANG K, WAN Y W, CHOI W T, et al. Combinatorial therapy discovery using mixed integer linear programming[J]. Bioinformatics, 2014,30(10):1456-1463.

[7] 吕琳媛. 复杂网络链路预测[J]. 电子科技大学学报, 2010(5):651-661.

[8] SID, REDNER. Networks: teasing out the missing links[J]. Nature, 2008, 453(7191):47-48.

[9] YU H, BRAUN P, YILDIRIM M A, et. al. High-quality binary protein interaction map of the yeast interactome network[J]. Science, 2008, 322(5898):104-110.

[10] 贾成龙. 基于网络模型和多源数据集成预测长非编码 RNA-疾病关联关系[D]. 西安:西安电子科技大学, 2017.

[11] 杨超,朱东华,衡晓帆,等. 基于语法树的 SAO 结构识别方法研究[J]. 图书情报工作, 2016,60(21):113-121.

[12] 杨超, 朱东华, 汪雪锋,等. 专利技术主题分析:基于 SAO 结构的 LDA 主题模型方法[J]. 图书情报工作, 2017,61(3):86-96.

[13] YOON J. Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis[J]. Scientometrics, 2013,94(1):313-331.

[14] 翟东升, 蔡文浩, 张杰,等. 基于图相似度的专利侵权检测方法研究[J]. 图书情报工作, 2018,62(5):97-105.

[15] 丁云轩, 闫雷. 数据挖掘软件 SemRep 的评价[J]. 中华医学图书情报杂志, 2008, 17(6):71-75.

[16] XU R, LI L, WANG Q. dRiskKB: a large-scale disease-disease risk relationship knowledge base constructed from biomedical text[J]. BMC bioinformatics, 2014, 15:105.

[17] XU K J, SONG J, ZHAO X M. The drug cocktail network[J]. BMC system biology, 2012, 6(S1): S5.

[18] CHEN X, REN B, CHEN M, et al. NLLSS: predicting synergistic drug combinations based on semi-supervised learning[J]. PLoS computational biology, 2016, 12(7): e1004975.

[19] JANSEN G, LEE A Y, EPP E, et al. Chemogenomic profiling predicts antifungal synergies[J]. Molecular systems biology, 2009, 5(1):338.

[20] 刘彦斌. 组合药物数据库构建及药物协同作用机理分析[D]. 杭州:浙江大学, 2011.

[21] RITAB A, ANJA T, JACOB B, et al. Species-specific activity of antibacterial drug combinations[J]. Nature, 2018, 559(7713): 259-263.

[22] HANLEY J A, MCNEIL B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve[J]. Radiology, 1982, 143(1): 29-36.

[23] 蒋智华. 艾滋病合并隐孢子虫感染的研究进展[J]. 中国艾滋病性病, 2014, 20(12):955-958.

[24] CRUZ P D, PRIDEAUX L, WAGNER J, et al. Characterization of the gastrointestinal microbiota in health and inflammatory bowel disease[J]. Inflammatory bowel diseases, 2012(2):372-390.

[25] FESTEN E A M, GOYETTE P, GREEN T, et al. A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2, TAGAP, and PUS10 as shared risk loci for crohn's disease and

celiac disease[J]. PLoS genetics, 2011, 7(1): e1001283.

作者贡献说明:

雷鸣:撰写和修改论文;

夏梦鸽:案例研究,数据分析,论文撰写及修改;

汪雪锋:论文选题和写作思路提出,论文定稿;

刘佳:整理文献,修改论文。

Research on Drug Combination Recommendation Based on Link Prediction for Concurrent Diseases Treatment

Lei Ming[1] Xia Mengge[1] Wang Xuefeng[1] Liu Jia[2]

[1] School of Management and Economics, Beijing Institute of Technology, Beijing 100081

[2] State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing 100024

Abstract: [**Purpose/significance**] Compared with single drug, drug combination has many advantages in clinical treatment. But the growth of drug quantity brings difficulties to drug combination screening experiment. Therefore, it is of great significance to design an effective prediction method to recommend drug combination which is more likely to produce synergistic effect for pharmaceutical staff, so as to improve the screening efficiency. [**Method/process**] For the need of concurrent diseases treatment, proposed a drug combination recommendation model based on link prediction, and used the SAO semantic mining to identify the complications in medical literature. On this basis, we used the medical database to build the heterogeneous "disease-drug-target" network, and introduced link prediction to evaluate the similarity of drug action mechanism, and predicted which drug combinations were more likely to have synergistic effect. Based on the prediction results, recommended a combination of drugs for a certain disease or a pair of complications. [**Result/conclusion**] The empirical analysis of intestinal disease data verified the practicality and efficiency of the model.

Keywords: SAO semantic mining link prediction social network analysis drug combination recommendation disease complications